

Developing and Applying Grading Criteria

Trying to decide the relative merits of a piece of writing can lead to a tangle of problems. Given a set of student essays, instructors frequently disagree, often vehemently, with one another's assessments. Because we teachers have little opportunity to discuss grading practices with colleagues, we often develop personal criteria that can seem eccentric to others. In fact, the first half-hour of a paper-grading workshop can be demoralizing even to the most dedicated proponents of writing across the curriculum. What do teachers actually want when they ask students to write?

Answering this question is not easy. Professional writing teachers grant that the assessment of writing, like the assessment of any art, involves subjective judgments. But the situation is not entirely relative either, for objective standards for good writing can be formulated, and readers with different tastes can be trained to assess writing samples with surprisingly high correlation. But the potential for wide disagreement about what constitutes good writing is a factor with which both students and teachers must contend.

The Problem of Criteria

The extent of this disagreement was illustrated by Paul Diederich (1974) in one of the most famous experiments in composition research. Diederich collected three hundred essays written by first-year students at three different universities and had them graded

by fifty-three professional persons in six different occupational fields. He asked each reader to place the essays in nine different piles in order of "general merit" and to write a brief comment explaining what he or she liked and disliked about each essay. Diederich reported these results: "Out of the 300 essays graded, 101 received every grade from 1–9; 94 percent received either seven, eight, or nine different grades; and no essay received less than five different grades" (p. 6).

Diederich discovered, however, some order in this chaos. Through factor analysis, he identified five subgroups of readers who correlated highly with one another but not with readers in other subgroups. By analyzing the comments on the papers, Diederich concluded that each subgroup was consistently giving predominant weight to a single criterion of writing. Sixteen readers were putting main emphasis on quality of ideas; thirteen on sentence structure, usage, spelling, and punctuation; nine on organization and development; nine on creative wording or phrasing; and seven on liveliness or committed voice, a factor Diederich labeled "flavor and personality." (Diederich counted one reader in two categories; hence these numbers add up to fifty-four rather than fifty-three; see his book, pp. 6–10, for details.)

Diederich's research enabled him to develop procedures through which a diverse group of readers could be trained to increase the correlation of their grading. By setting descriptions for high, middle, and low achievement in each of his five criterion areas—ideas, organization, sentence structure, wording, and flavor—Diederich was able to train readers to balance their assessments over the five criteria. Since then, numerous researchers have refined or refocused Diederich's criteria and have developed successful strategies for training readers as evaluators (see, particularly, Cooper and Odell, 1977, and White, 1992, 1994). Many of these strategies have classroom applications also, for training students as evaluators of writing greatly improves their ability to give high-quality advice in peer review workshops.

Providing Criteria for Students

Even though readers can be trained to apply uniform criteria to student essays, these criteria often vary from discipline to discipline (and from teacher to teacher), a phenomenon that often confuses students. Not only do styles vary widely across the disciplines, but there are also fundamental differences in the way arguments are structured and elaborated—a problem students feel acutely as they move through their general education courses.

To make matters more confusing for students, different teachers within the same discipline often value different kinds of writing. Some teachers, as we have seen in Chapter Three, want students to sound like professionals in the field. Others assign narratives, personal reflections, and other alternative assignments calling for voices other than the apprentice academic.

Because of such variety of expectations, instructors should describe their criteria for judging writing and, whenever possible, provide samples of successful student papers from previous classes.

Developing Criteria and Grading Scales

Criteria for writing are usually presented to students in one of two ways: analytically or holistically. The analytic method gives separate scores for each criterion—for example, ideas, ten points; organization, ten points; sentence structure, five points—whereas the holistic method gives one score that reflects the reader's overall impression of the paper, considering all criteria at once. Many instructors prefer analytic scales because the breakdown of the grade into components, when combined with the instructor's written comments, conveys detailed information about the teacher's judgment of the essay. Some people object philosophically to analytic scoring, however, on the grounds that writing cannot be analyzed into component parts. Can ideas really be separated from organization or clarity of expression from clarity of thought? Such people prefer holistic evaluation, which does not suggest that writing is a mixture of separable elements. Also, holistic grading is faster and so is often preferable when one's main concern is rapidity of assessment rather than precision of feedback.

Both analytic and holistic scoring methods can also be classified two ways: general description methods and primary trait methods. Proponents of general description argue that criteria for writing can be stated in a general or universal way (good organization, graceful sentence structure, and so forth). Proponents of the primary trait method, however, argue that criteria must be stated specifically in terms of the given writing task. For example, the criteria for a history paper detailing the origins of the electoral college would differ from those of a political science paper arguing that the electoral college should be abolished. A primary trait scale for the history paper might include criteria like these:

Does the writer make effective use of primary sources?

Does the essay explore the alternatives to the electoral college discussed at the constitutional convention?

In contrast, a primary trait scale for the political science paper might include these criteria:

- Does the writer predict the consequences of abolishing the electoral college using acceptable empirical data?
- Does the writer anticipate objections to these predictions and adequately respond to them?

Thus, a primary trait scale uses grading criteria keyed directly to the assignment. (Examples of different kinds of grading scales will appear later in this chapter.)

Developing Analytic Scales

Exhibit 15.1 illustrates a simple analytic scale using general description methods. Analytic scales normally list three or more criteria, almost always including quality of ideas, organization, and sentence structure. Many analytic scales are elaborate, with numerous additional categories and subcategories. Some analytic scales are dichotomous, meaning that the reader simply checks off “yes” or “no,” depending on the presence or absence of certain features of the writing:

Is there a thesis statement? Yes _____ No _____

Other scales ask the reader to rate each feature of the writing along a number sequence:

Quality of thesis statement:

Low			Middle			High	
1	2	3	4	5	6	7	8

Many analytic scales weigh some criteria more heavily than others, depending on what the instructor wishes to emphasize. Thus, you might allot twenty-five points for ideas, fifteen points for organization, and ten points for sentence structure. But if you are particularly annoyed by careless spelling errors, you might give ten bonus points to papers with no misspelled words and deduct ten points for having more than, say, five misspelled words. Exhibits 15.2 and 15.3 illustrate analytic scales using primary trait criteria. Exhibit 15.2 is a scoring guide developed by an English professor for an assignment on Conrad’s *The Secret Sharer*. The professor gives the scoring guide to students at the time she passes out the assignment. The scoring guide thus reinforces key features she expects in students’ essays and serves as a checklist during peer review. Exhibit 15.3 is a scoring guide used by finance professor Dean Drenk to pro-

Exhibit 15.1. Simple Analytic Scale (General Description Method).

Scoring Guide for Essays**Quality of Ideas (____ points)**

Range and depth of argument; logic of argument; quality of research or original thought; appropriate sense of complexity of the topic; appropriate awareness of opposing views.

Organization and Development (____ points)

Effective title; clarity of thesis statement; logical and clear arrangement of ideas; effective use of transitions; unity and coherence of paragraphs; good development of ideas through supporting details and evidence.

Clarity and Style (____ points)

Ease of readability; appropriate voice, tone and style for assignment; clarity of sentence structure; gracefulness of sentence structure; appropriate variety and maturity of sentence structure.

Sentence Structure and Mechanics (____ points)

Grammatically correct sentences; absence of comma splices, run-ons, fragments; absence of usage and grammatical errors; accurate spelling; careful proofreading; attractive and appropriate manuscript form.

vide feedback on his thesis support microthemes in finance (see Chapter Five, pages 74–75). His scoring guide can be easily adapted to the needs of professors in other disciplines.

Developing Holistic Scales

Samples of holistic scales are shown in Exhibits 15.4 and 15.5. Exhibit 15.4 is a holistic scale for summary-writing assignments. Exhibit 15.5 is a holistic scale for grading physics microthemes. Holistic scoring depends on a reader's all-at-once assessment of a paper based on one attentive but quick reading. Research suggests that the correlation between readers actually increases if readers read quickly, trusting the reliability of their first impressions (White, 1994). Thus, holistic scales work best in conjunction with rapid grading and "models feedback" (see Chapter Thirteen, page 236; see also Rogers, 1995, for a discussion of holistic scoring in a chemistry course).

Conducting a Departmental Norming Session _____

A good way to improve one's grading practices is to join a conversation with colleagues about what constitutes excellent, good,

Exhibit 15.2. Analytic Scale (Primary Trait Method).

Scoring Guide for Assignment on *The Secret Sharer*

Your essay is supposed to provide a supported answer to the following question:

How has the experience with Leggatt changed the captain so that what he is at the end of the story is different from what he was at the beginning?

In order to do well on this paper, you need to do these things:

1. Have your own clear answer to this question.
2. Support your answer with strong arguments and textual details.
3. Make your essay clear enough for a reader to understand with one reading.

Criterion 1. Does your essay have a thesis statement at the end of the first paragraph that answers the question regarding changes in the captain?

no thesis or unclear thesis			clear thesis		
2	4	6	8	10	

Criterion 2. Is your thesis supported with strong argumentation and use of significant details taken from the story?

weak argument and/or lack of details as support			strong argument and good details as support		
2	4	6	8	10	

Criterion 3. Is your paper easy for a reader to follow?

Paragraphing and transitions	2	4	6	8	10
Clear Sentences	2	4	6	8	10
Accurate mechanics: grammar, spelling, punctuation, neatness	2	4	6	8	10

Source: Dolores Johnson

satisfactory, and poor papers. A surefire way to stimulate such conversation is to "staff-grade" with colleagues a set of essays written in response to an assignment within your discipline. One participant selects in advance four or five essays that seem to span the range of quality from excellent to poor, duplicates them for the department, and uses them to initiate discussion. In developing criteria, instructors are advised to use a number scale that does not translate directly into letter grades. A six-point scale ranging from 6 (best) to 1 (worst) is most common. Using a numerical scale tem-

Exhibit 15.3. Primary Trait Scoring Guide for Thesis Support Essays in Finance.

Grading Criteria	
Support of Theses	
A. Clarity of support:	_____
B. Logic (relationship of support to thesis):	_____
C. Sources of support	
1. Quantity	_____
2. Quality	_____
Total microtheme grade	_____
Specific Features of Your Microtheme	
_____	Grammmatical errors are numerous enough to interfere with understanding your response.
_____	The organization of your response is not clear.
_____	The logic of your support is confusing or does not make sense.
_____	Your conclusions are not warranted by your support.
_____	Your support is too imprecise or too general.

Source: Bean, Drenk, and Lee, 1982, p. 32.

porarily suspends the additional problem of variable standards for letter grades. Thus, a "hard grader" and an "easy grader" might agree that a particular essay rates a 4 on a six-point scale but disagree on how to translate that 4 into a letter grade. The hard instructor might give it a C+ and the easy instructor a B. Since standards for letter grades are a different issue from standards for ranking several pieces of writing, problems of devising criteria for writing are simplified if we separate the two issues, at least initially.

After an initial norming session in which department members reach agreement on the sample papers and develop criteria for each gradation on their scoring scale, members break into pairs to staff-grade the set of essays. Each essay is read independently by two readers, who meet periodically to compare scores and discuss discrepant grades. On a six-point scale, instructors should aim to come within one point of each other's scores. Differences of two or more points indicate a wide divergence of criteria. A departmental norming session every year or so can increase instructors' communal

Exhibit 15.4. Holistic Scale for Grading Article Summaries.

A summary should be directed toward imagined readers who have not read the article being summarized. The purpose of the summary is to give these persons a clear overview of the article's main points. The criteria for a summary are (1) accuracy of content, (2) comprehensiveness and balance, and (3) clarity, readability, and grammatical correctness.

- 6** A 6 summary meets all the criteria. The writer understands the article thoroughly. The main points in the article appear in the summary with all main points proportionately developed (that is, the writer does not spend excessive time on one main point while neglecting other main points). The summary should be as comprehensive as possible and should read smoothly, with appropriate transitions between ideas. Sentences should be clear, without vagueness or ambiguity and without grammatical or mechanical errors.
- 5** A 5 summary should still be very good, but it can be weaker than a 6 summary in one area. It may have excellent accuracy and balance, but show occasional problems in sentence structure or correctness. Or it may be clearly written but be somewhat unbalanced or less comprehensive than a 6 summary or show a minor misunderstanding of the article.
- 4** A score of 4 means "good but not excellent." Typically, a 4 summary will reveal a generally accurate reading of the article, but it will be noticeably weaker in the quality of writing. Or it may be well written but cover only part of the essay.
- 3** A 3 summary must have strength in at least one area of competence, and it should still be good enough to convince the grader that the writer has understood the article fairly well. However, a 3 summary typically is not written well enough to convey an understanding of the article to someone who has not already read it. Typically, the sentence structure of a 3 summary is not sophisticated enough to convey the sense of hierarchy and subordination found in the essay.
- 2** A 2 summary is weak in all areas of competence, either because it is so poorly written that the reader cannot understand the content or because the content is inaccurate or seriously disorganized. However, a 2 essay convinces the grader that the writer has read the essay and is struggling to understand it.
- 1** A 1 summary fails to meet any of the areas of competence.

confidence in their grading practices. For more detailed descriptions of this procedure, along with sample student essays and reader-developed scoring criteria, see White (1992). See also Bateman's discussion of scoring a set of sociology essays dealing with ethnocentrism (1990, pp. 110–116).

Determining Grades

Assigning a letter grade to a piece of writing always poses a dilemma, and I can offer no easy advice. Teachers who use analytic scales often add up each student's total score, rank the papers, and trans-

Exhibit 15.5. Holistic Scale for Grading Physics Microthemes.

- 6, 5** Microthemes in the category will show a confident understanding of the physics concepts and will explain those concepts clearly to the intended audience. A 6 theme will be clearly written throughout; will contain almost no errors in spelling, punctuation, or grammar; and will have enough development to provide a truly helpful explanation to learners. A 5 theme will still be successful in teaching the physics concepts to the intended audience but may have more errors or somewhat less development than a 6. The key to microthemes in the 6, 5 category is that they must show a correct understanding of the physics and explain the concept clearly to a new learner.
- 4, 3** Microthemes in this category will reveal to the instructor that the writer probably understands the physics concepts, but lack of clarity in the writing or lack of fully developed explanations means that the microtheme would not teach the concept to new learners. Microthemes in the 4, 3 category are usually "you know what I mean" essays: someone who already understands the concepts can tell that the writer probably does, too, but someone who does not already understand the concepts would not learn anything from the explanation. This category is also appropriate for clearly written essays that have minor misunderstandings of the physics concepts or for accurate essays full of sentence-level errors.
- 2, 1** These microthemes will be unsuccessful either because the writer fails to understand the physics concepts, because the number of errors is so high that the instructor cannot determine how much the writer understands, or because the explanations lack even minimum development. Give a score of 2 or 1 if the writer misunderstands the physics, even if the essay is otherwise well written. Also give a score of 2 or 1 to essays so poorly written that the reader can't understand them.

late scores into letter grades by establishing a curve or by setting point ranges for levels of grades. Other teachers, using a more holistic method, try to develop an interior sense of what an A, B, C, or D essay looks like. If possible, it is best to read through a set of papers quickly before marking them and assigning grades, trying to get a feel for the range of responses and sizing up what the best papers are like. In grading essay exams or short papers, many teachers develop schemes for not knowing who the authors are until the papers are graded. (One method is to have students use their social security numbers rather than names; another is to have students put their names on the back of the last page.) Not knowing who wrote which essay eliminates any halo effect that might bias the grade.

To avoid grading on the curve, some teachers like to establish criteria for grading that are as objective and as consistent as possible. Although this is no easy task, the following explanation, written by Cornell University English professor Harry Shaw (1984), shows how one professor makes his decision. It is as good a guide as any I know.

How I Assign Letter Grades

In grading "thesis papers" . . . I ask myself the following set of questions:

1. Does the paper have a thesis?
2. Does the thesis address itself to an appropriate question or topic?
3. Is the paper free from long stretches of quotations and summaries that exist only for their own sakes and remain unanalyzed?
4. Can the writer produce complete sentences?
5. Is the paper free from basic grammatical errors?

If the answer to any of these questions is "no," I give the paper some kind of C. If the answer to most of the questions is "no," its grade will be even lower.

For papers which have emerged unscathed thus far, I add the following questions:

6. How thoughtful is the paper? Does it show real originality?
7. How adequate is the thesis? Does it respond to its question or topic in a full and interesting way? Does it have an appropriate degree of complexity?
8. How well organized is the paper? Does it stick to the point? Does every paragraph contain a clear topic sentence? If not, is another kind of organizing principle at work? Are the transitions well made? Does it have a real conclusion, not simply a stopping place?
9. Is the style efficient, not wordy or unclear?
10. Does the writing betray any special elegance?
11. Above all, can I hear a lively, intelligent, interesting human voice speaking to me (or to another audience, if that's what the writer intends) as I read the paper?

Depending on my answers to such questions, I give the paper some kind of A or some kind of B [pp. 149–150].

Conclusion: Expecting Excellence

When students know an instructor's criteria for assigning grades—and when they have the opportunity to help one another apply these criteria to works in progress—the quality of their final products will improve gratifyingly. It is satisfying indeed to see how well many undergraduates can write when they are engaged in their projects and follow the stages of the writing process through multiple drafts and peer reviews. By setting high standards, by encouraging multiple drafts, by refusing to be the first human being to read a student's paper—in short, by expecting excellence—instructors can feel justified in applying rigorous criteria.

But it is important too that students never think of their writing as "finished." In the best of all worlds, students would be allowed to rewrite a paper if they wished to improve it further. The presence of grades should never override the more important emphasis on revision and improvement.

The point, then, of assigning writing across the curriculum is to engage students in the process of inquiry and active learning. Although one of our goals is to improve students' communication skills, writing is more than communication; it is a means of learning, thinking, discovering, and seeing. When teachers give students good problems to think about—and involve them actively in the process of solving these problems—they are deepening students' engagement with the subject matter and promoting their intellectual growth. By adding well-designed writing assignments to a course, teachers give students continued practice in critical thinking. Teachers know when their approach is working: the performance of their students improves.